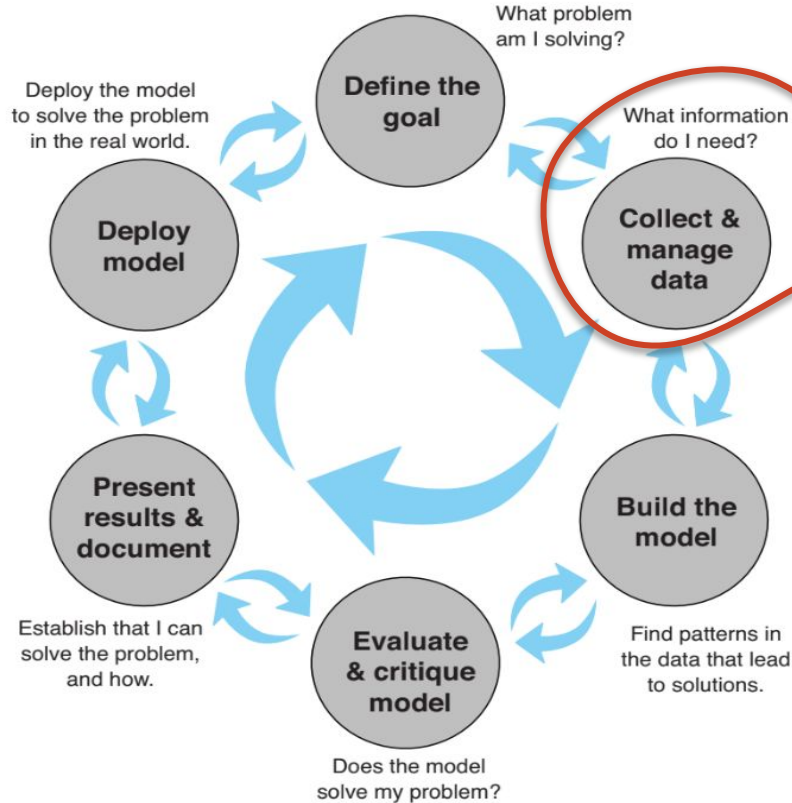


Laboratorio de Datos

Clasificación - KNN



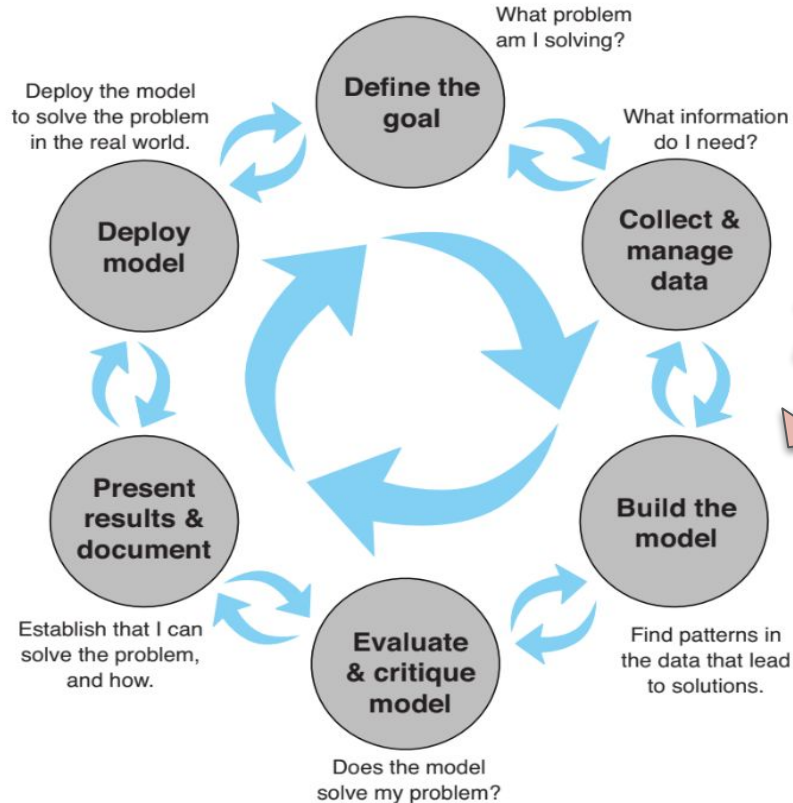
Recorrido de la materia (hasta ahora)



1º parte de la materia

- ✓ Lenguaje de programación (Python)
- ✓ Modelado conceptual de los datos (DER)
- ✓ Representación de los datos (modelo relacional)
- ✓ Formas de consultar los datos (AR/SQL)
- ✓ Recomendaciones para el diseño (Normalización)
- ✓ Calidad de datos
- ✓ Leyes acerca de la Protección de Datos

Recorrido de la materia (hasta ahora)



- ✓ *Visualización y Exploración de los datos*
- ✓ *Intro a Modelado: Clasificación y Regresión*
- ✓ *Clasificación: Árboles de decisión*
- ✓ *Regresión: Regresión Lineal, KNN*

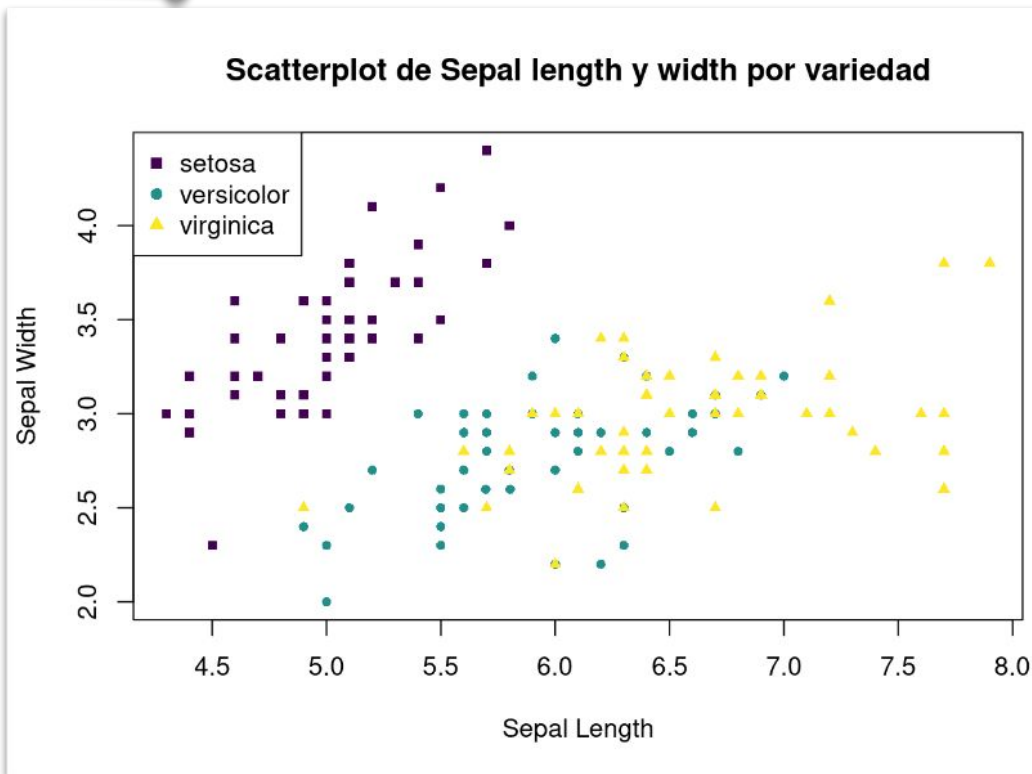
Clasificación con K Nearest Neighbors (KNN)

... por Manuela Cerdeiro (y modificaciones de P. Turjanski)

Clasificación con *K* Nearest Neighbors (KNN)

Es similar a cuando lo usamos para regresión:

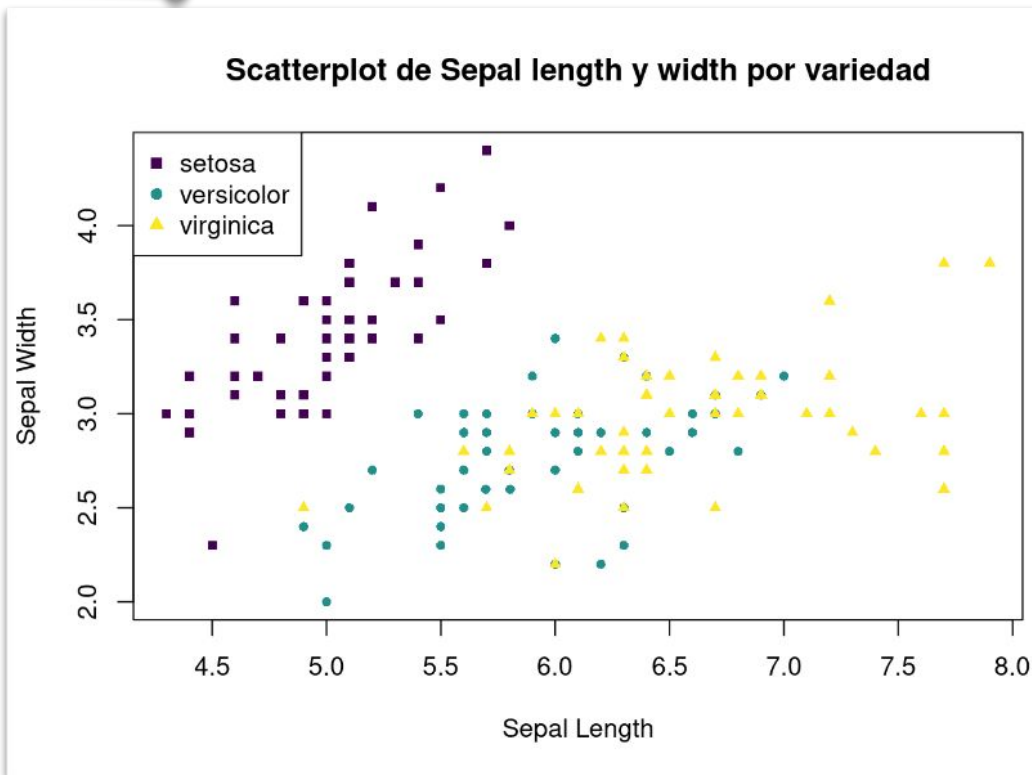
1. *Datos de entrenamiento
(variables predictoras)*



Clasificación con *K* Nearest Neighbors (KNN)

Es similar a cuando lo usamos para regresión:

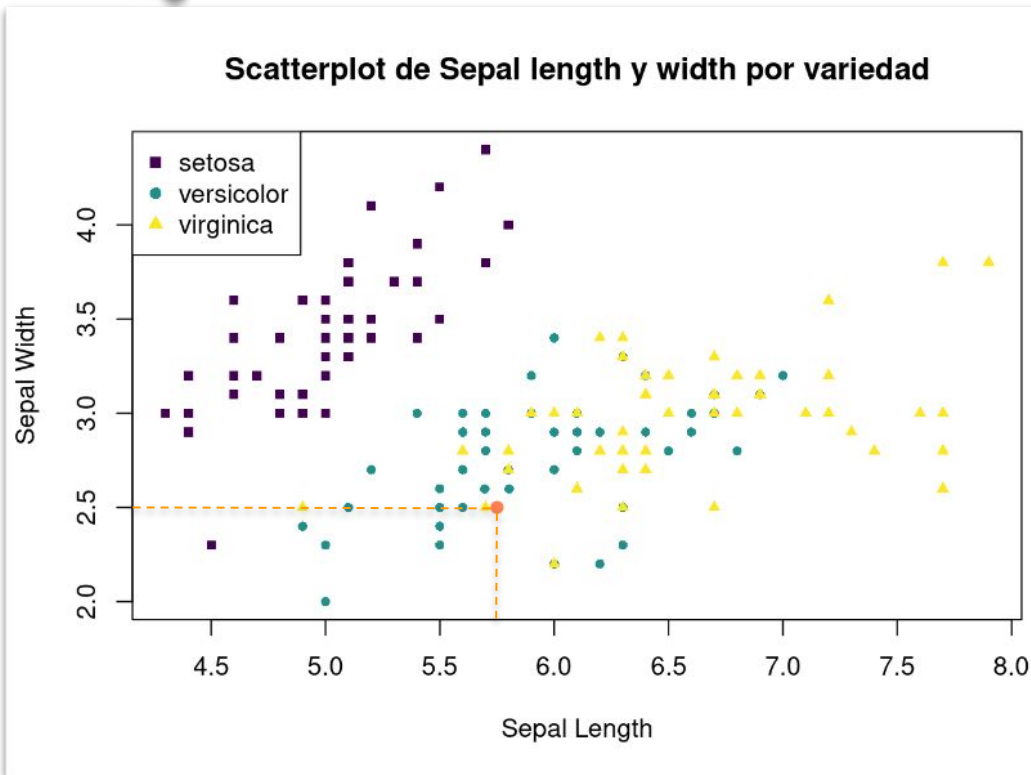
1. Datos de entrenamiento
(variables predictoras)
2. Definimos distancia: Euclídea



Clasificación con K Nearest Neighbors (KNN)

Es similar a cuando lo usamos para regresión:

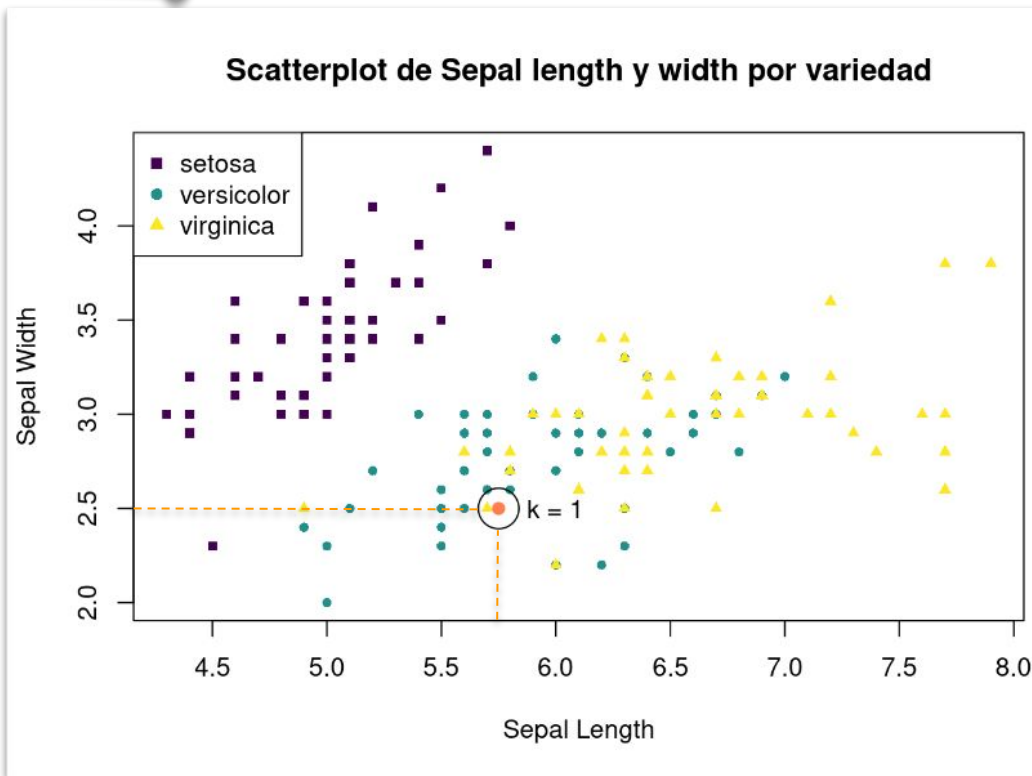
1. Datos de entrenamiento
(variables predictoras)
2. Definimos distancia: Euclídea
3. Recibimos una nueva instancia



Clasificación con *K* Nearest Neighbors (KNN)

Es similar a cuando lo usamos para regresión:

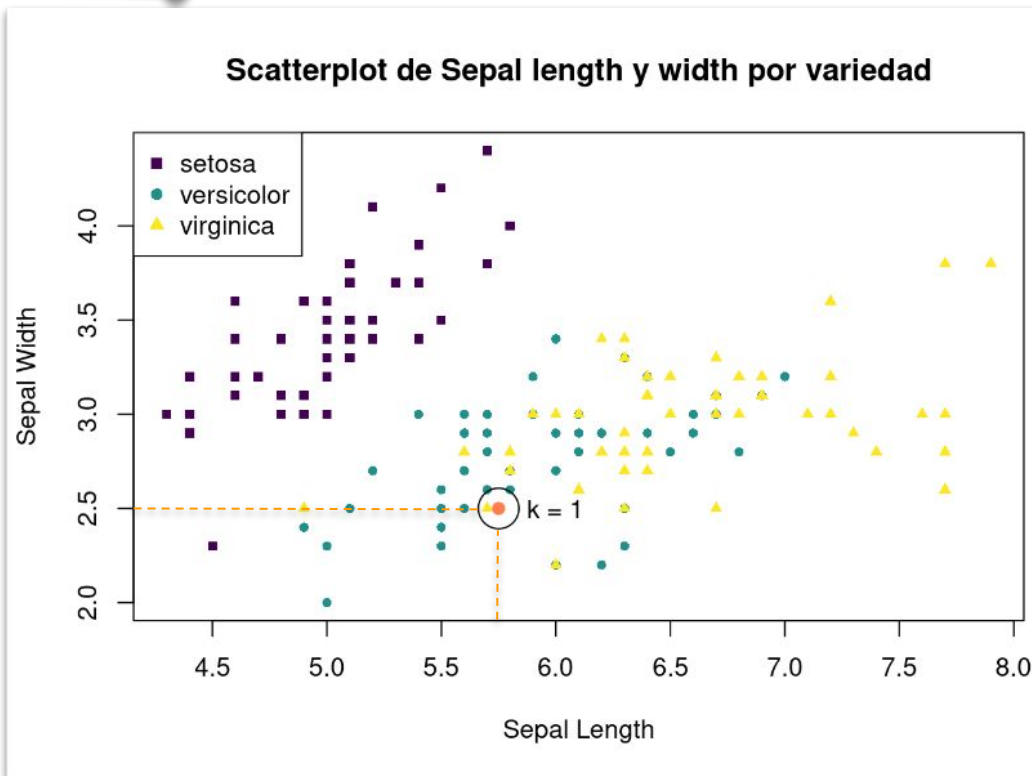
1. *Datos de entrenamiento (variables predictoras)*
2. *Definimos distancia: Euclídea*
3. *Recibimos una nueva instancia*
4. *k vecinos más cercanos (distancia Euclídea)*



Clasificación con K Nearest Neighbors (KNN)

Es similar a cuando lo usamos para regresión:

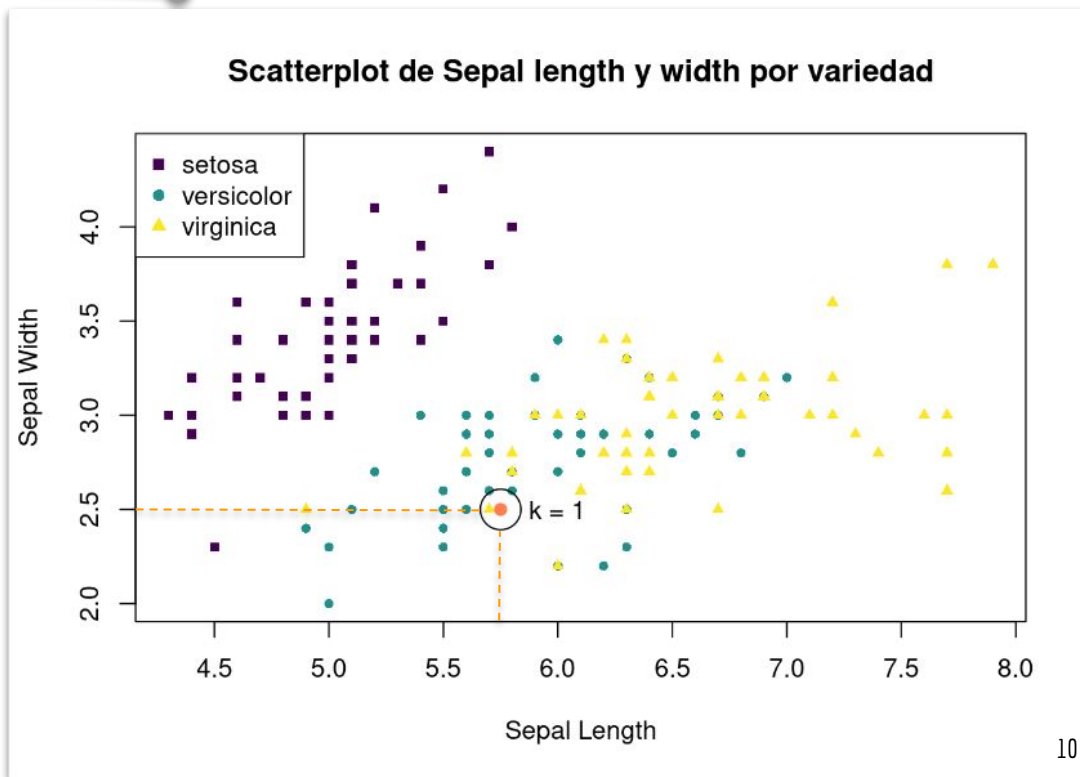
1. Datos de entrenamiento
(variables predictoras)
2. Definimos distancia: Euclídea
3. Recibimos una nueva instancia
4. k vecinos más cercanos (distancia Euclídea)
5. ¿Clase de los k vecinos?
(variable a predecir)



Clasificación con K Nearest Neighbors (KNN)

Es similar a cuando lo usamos para regresión:

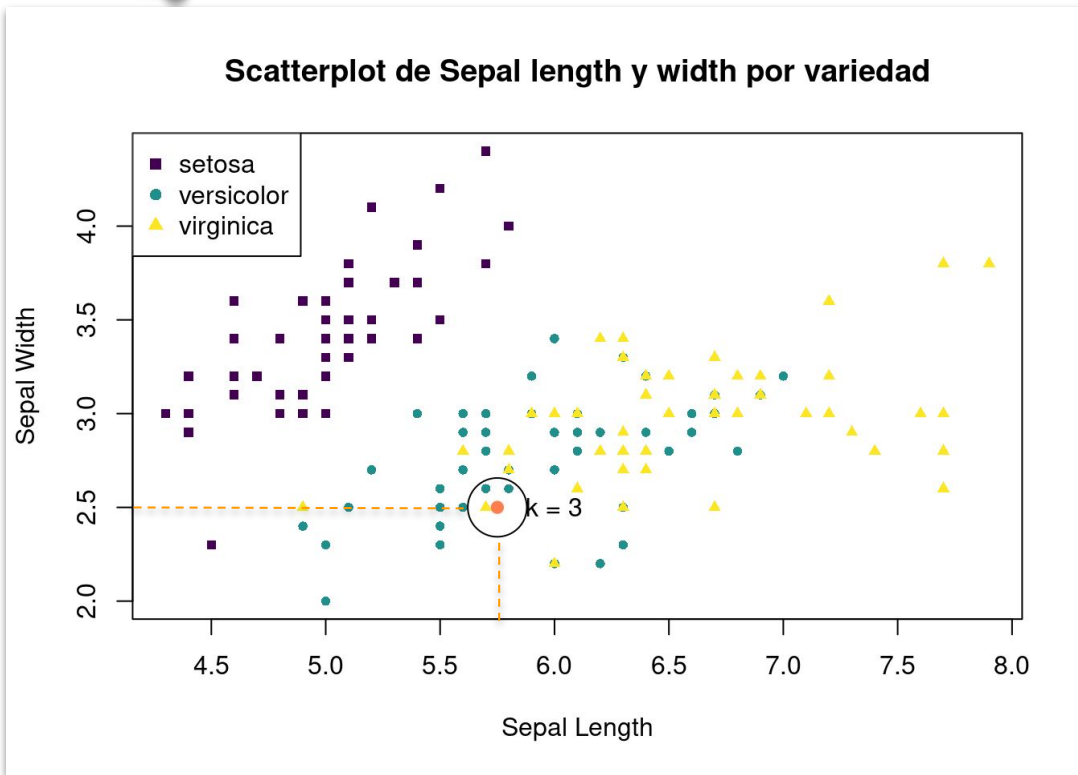
1. Datos de entrenamiento
(variables predictoras)
2. Definimos distancia: Euclídea
3. Recibimos una nueva instancia
4. k vecinos más cercanos (distancia Euclídea)
5. ¿Clase de los k vecinos?
(variable a predecir)
6. Elegimos la clase mayoritaria



Clasificación con K Nearest Neighbors (KNN)

Es similar a cuando lo usamos para regresión:

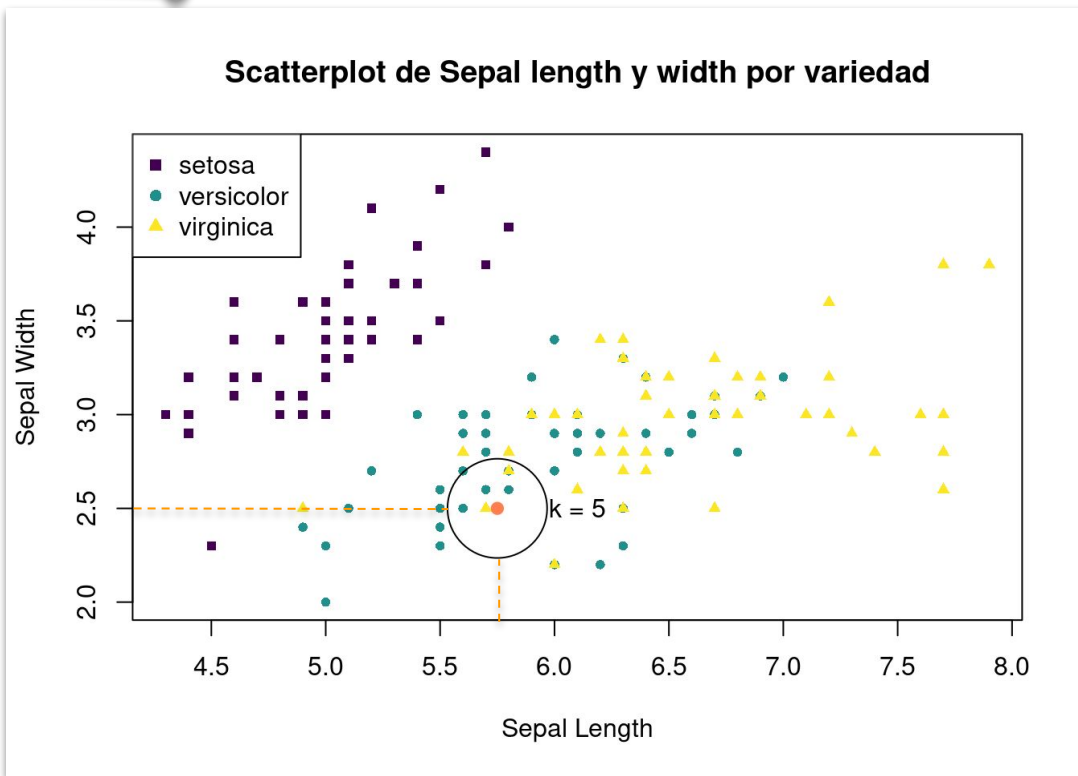
1. Datos de entrenamiento
(variables predictoras)
2. Definimos distancia: Euclídea
3. Recibimos una nueva instancia
4. k vecinos más cercanos (distancia Euclídea)
5. ¿Clase de los k vecinos?
(variable a predecir)
6. Elegimos la clase mayoritaria



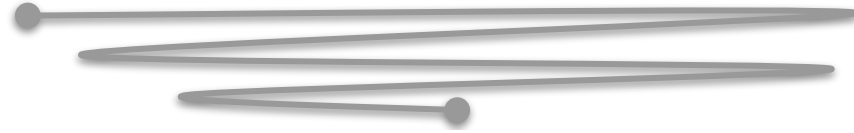
Clasificación con K Nearest Neighbors (KNN)

Es similar a cuando lo usamos para regresión:

1. Datos de entrenamiento
(variables predictoras)
2. Definimos distancia: Euclídea
3. Recibimos una nueva instancia
4. k vecinos más cercanos (distancia Euclídea)
5. ¿Clase de los k vecinos?
(variable a predecir)
6. Elegimos la clase mayoritaria



Ejemplo con Dataset Iris



Variables predictoras: Sépalo/Pétalo, Alto/Ancho (4 variables)

Variable a predecir : Target (tipo de flor)

Ejemplo con Dataset Iris



Variables predictoras: Sépalo/Pétalo, Alto/Ancho (4 variables)

Variable a predecir : Target (tipo de flor)

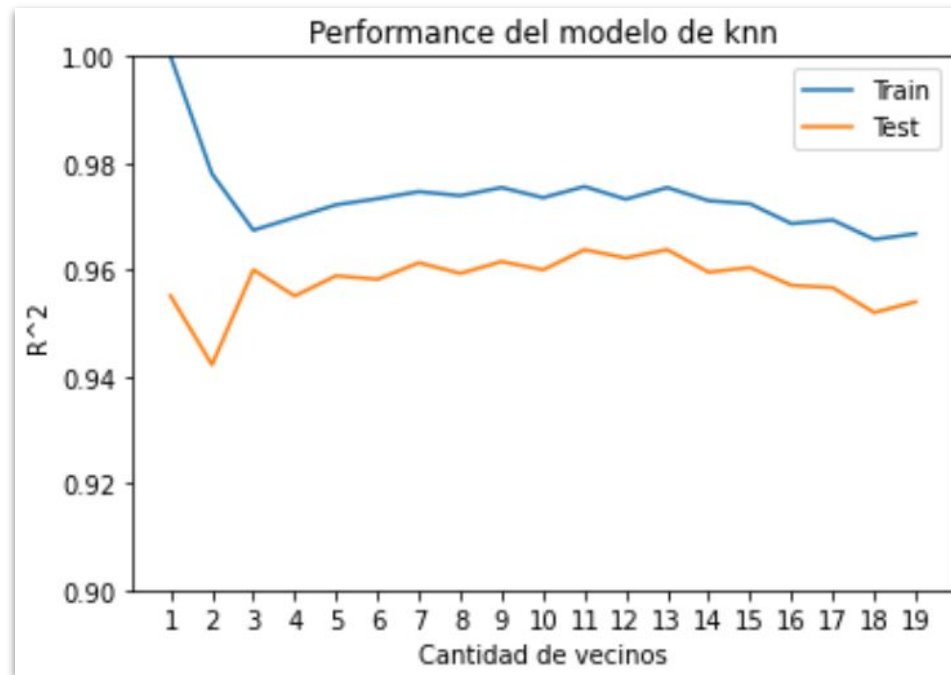
Estrategias:

1. Usamos todo el dataset
2. Usamos Train/Test
3. Repetimos con distintos Train/Test

Ejemplo con Dataset Iris



¿Cuál es el mejor valor de k ?

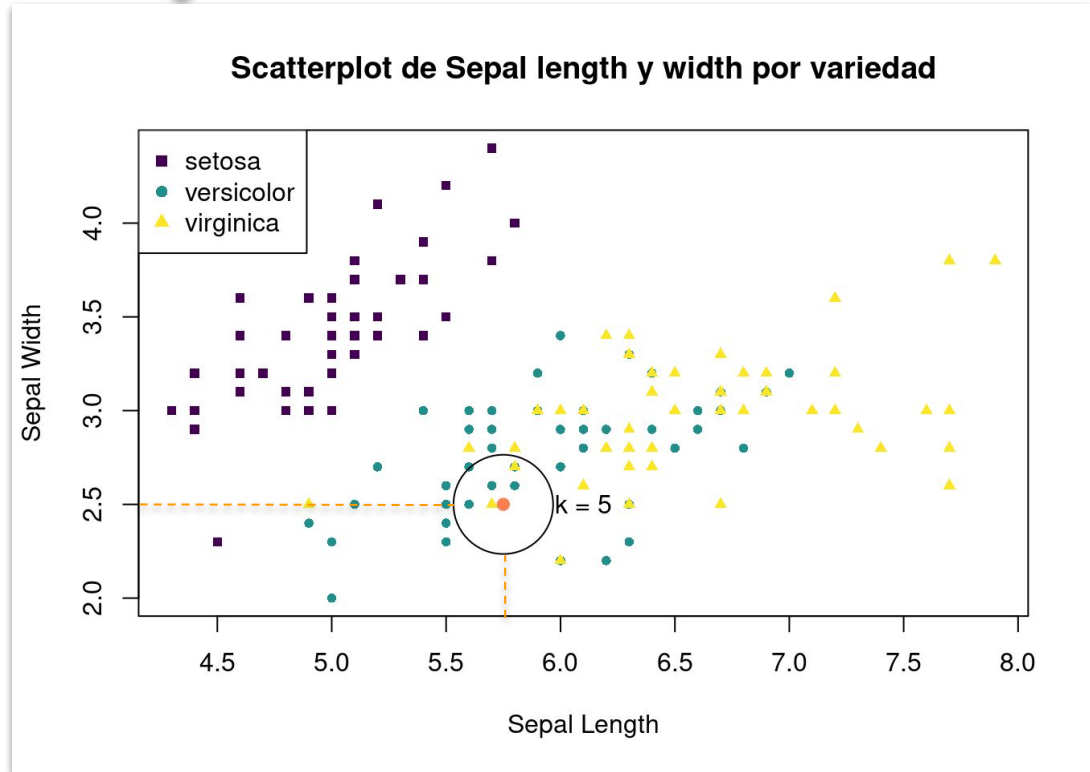


Advertencia

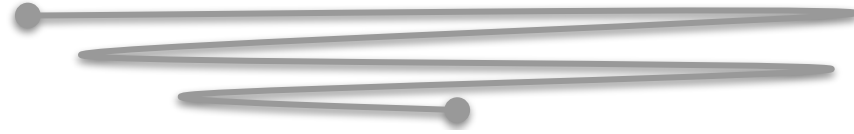
Es similar a cuando lo usamos para regresión:

1. Datos de entrenamiento
(variables predictoras)
2. Definimos distancia: **Euclídea**
3. Recibimos una nueva instancia
4. k vecinos más cercanos (distancia Euclídea)
5. ¿Clase de los k vecinos?
(variable a predecir)
6. Elegimos la clase mayoritaria

Puede ser un problema → estandarizar



Cierre



1. Presentamos el modelo KNN, pero esta vez adaptado para clasificación
2. Evaluamos su performance
3. Varía con k y suele ser peor en el conjunto de test (comparado con el train)